Journal of Biomedical Informatics 80 (2018) 43–51

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Augmentation of the think aloud method with users' perspectives for the selection of a picture archiving and communication system

Misagh Zahiri Esfahani[a], Reza Khajouei[b,c,*], Mohammad Reza Baneshi[d]

[a] Health Management and Economics Research Center, Iran University of Medical Sciences, Tehran, Iran
[b] Medical Informatics Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran
[c] Department of Health Information Sciences, Faculty of Management and Medical Information Sciences, Kerman University of Medical Sciences, Kerman, Iran
[d] Modeling in Health Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

ARTICLE INFO

Keywords:
Picture archiving and communication system
User-computer interface
Radiology information system
Think aloud
Usability evaluation
User interaction

ABSTRACT

Objectives: Users attitude toward a picture archiving and communication system (PACS) and their interaction with this system are among the most important factors that influence its acceptance. This study aimed to augment the user's interaction with the user's perspective to select a usable PACS among three systems available on the market.

Methods: We augmented the think aloud (TA) usability evaluation method with the Post-Study System Usability Questionnaire (PSSUQ) to compare user interaction problems of three PACS user interfaces. Four radiologists and four internist physicians participated in this study. Usability characteristics including efficiency, effectiveness, learnability, error, and satisfaction were used to assess the usability of each PACS.

Results: There was a significant difference in efficiency ($p = 0.01$), effectiveness ($p = 0.005$), learnability ($p = 0.001$), and satisfaction ($p = 0.009$). However, no significant difference in the number of errors ($p = 0.18$), mouse clicks and keystrokes ($p = 0.12$), and the number of usability problems ($p = 0.6$) were observed among the three PACS systems studied.

Conclusions: This study showed that applying the proposed approach to augment TA with the user's perspective addresses almost all of the theoretical aspects of usability and can be employed to select the most usable PACS.

1. Introduction

The shift from hard copy film-based imaging to digital imaging has significantly changed the workflow in radiology departments and medical institutions. The picture archiving and communication system (PACS) is one of the most important medical imaging technologies that have contributed to digital radiography [1,2]. PACS systems use an electronic archive for short- and long-term storage, retrieval, and management of medical images, a secure network for distribution, and workstations or mobile devices for the presentation of medical images produced by various medical hardware modalities, such as X-rays, CT scans, MRIs, and ultrasound machines [3].

Limited financial resources are the major challenge for the selection, installation, and maintenance of PACS systems in healthcare organizations [4,5]. While PACS systems are expensive, they are among the most important medical applications and are capable of bringing high returns on investment [6]. The selection and implementation of a PACS also faces other challenges and obstacles, including difficulty in selecting an appropriate PACS due to the lack of awareness of the best selection criteria [7,8]. Currently, the selection is done through a multidimensional comparison, such as price and technical functionalities between the commercially available PACS systems developed by different vendors. Although various aspects of the systems are reviewed via this approach, some necessary concerns might remain unaddressed. It is important to clarify whether the selected software is easy to use since the software-user interaction plays an important role in the application performance [9,10]. Although having different functionalities might seem promising and lead to a higher rank in comparison, users may find them confusing, difficult, and sometimes even impossible to use. User-software interaction that fails to meet the users' needs reduces the efficiency, satisfaction, productivity, and accuracy [11,12]. Several studies have reported rejected or withdrawn health information systems due to difficulty of use [13–15]. Therefore, appropriate considerations must be given to these perspectives in order to enhance the successful adoption of a PACS [16].

Previous studies on the usability of PACS systems [17–24] identified

* Corresponding author at: Medical Informatics Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Haft-Bagh Highway, P.O. Box: 7616911320, Kerman, Iran.
E-mail address: r.khajouei@kmu.ac.ir (R. Khajouei).

https://doi.org/10.1016/j.jbi.2018.02.017
Received 25 July 2017; Received in revised form 15 February 2018; Accepted 26 February 2018
Available online 28 February 2018
1532-0464/

problems that can make the interaction time consuming, causing delays in tasks, dissatisfaction, and frustration, preventing users from enjoying all of the benefits and functionalities of the system, as well as leading to more errors and difficulties in performing clinical analyses [25]. Furthermore, a recent review article showed the lack of studies on usability evaluations of PACS systems using formal evaluation methods involving the final users [25]. Some have either evaluated the user interfaces of a single PACS [19,21,22,24], or as a subsidiary part [23,26], but to the best of our knowledge only a few studies have specifically focused on and addressed the evaluation of the user interface of different PACS systems [27,28]. They have investigated the user's perspectives to compare PACS systems with no intention of applying this information for selection. Jorritsma et al. [18] investigated the user's perspectives and interactions in a comparative study for the selection of one of four PACS systems. The study was conducted on radiologists as a group of PACS users and used a webcam for collecting data in testing sessions and finally concentrated on the analysis of satisfaction and efficiency characteristics.

Recruiting different groups of users, employing a specialized tool as well as investigating other usability characteristics such as effectiveness, learnability, and errors when real users interact with the system can add to the existing knowledge and provide more insights into the design of a PACS user interface.

At the time of this study, Kerman University of Medical Sciences was in the process of selecting and purchasing a PACS. Three demonstration PACS systems were provisionally installed in three different medical centers for the purpose of comparison and evaluation. Since the users' interaction with a PACS and their view after the first interaction with the system have a major impact on the success and adoption of the PACS, this study was designed to compare users' interaction problems as well as perspectives about the three PACS systems in terms of different usability characteristics.

2. Methods

2.1. Design and procedure

Three common PACS applications, hereafter called A, B, and C, from Iranian PACS vendors were evaluated in this comparative study. One vendor supplied an imported PACS and the other two provided their own developed software.

The think aloud (TA) method and the Post-Study System Usability Questionnaire (PSSUQ) were used to study the user's interaction and perspective, respectively. We augmented these two methods to be able to measure all the characteristics contributing to a usability test.

TA is the gold standard of usability evaluation [29], concentrating on a user's cognition while interacting with a system. In this method, users are asked to verbalize their feelings, thoughts, and whatever else comes to mind while performing tasks on a series of predetermined scenarios. The task examples should be as realistic as possible and representative of end-user performances in daily life situations. During the session, there should be full audiotaping and/or video recording of the participants and, if possible, video recording of the computer screens to document all important information. Usability problems are detected by evaluators from analyses of user behavior and expressions during interactions with the system [30,31]. The participants should be a sample of users representing the expected end users.

The PSSUQ consists of 19 items that were designed for immediate administration after usability testing [32]. The PSSUQ utilizes the 5-point Likert scale from "strongly agree" to "strongly disagree" and has two text fields for any comment and explanation by the participants. The PSSUQ was first translated into Persian and, to approve the cross-cultural comparison of translation, was re-translated into English by an expert and then its compliance with the original text was confirmed. The content validity of the PSSUQ was confirmed by one radiologist and three medical informatics experts. The reliability was determined using Cronbach's alpha ($\alpha = 0.96$).

To define the measurement criteria, a coding framework was developed according to five usability characteristics and based on the International Organization for Standardization (ISO) and Nielsen's definitions [33–35] to recognize the specific user-computer interaction problems in detail. According to the ISO, usability is defined as "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use." Nielsen put forward five usability attributes: learnability, efficiency, memorability, errors, and satisfaction [35]. Combining ISO and Nielsen usability attributes yields the following six criteria: efficiency, effectiveness, learnability, memorability, errors, and satisfaction. Since the participants in this study used each system only once and there was no need to remember the options for a next session, we did not consider memorability in our evaluation. The remaining five attributes composed our coding framework.

We used the TA method to measure effectiveness, learnability, errors, and efficiency characteristics, and the PSSUQ was used to measure satisfaction.

2.2. Participants

Since five to eight participants are considered sufficient to perform TA [36,37], we recruited eight participants from two user groups. The participants included four radiologists and four internist physicians from a university hospital in Kerman, Iran. Since users mostly interact with a PACS via PACS viewers, the focus of this usability study was on evaluating the users' interaction with the PACS viewers.

Three evaluators acted as facilitators of the testing sessions and analyzed the results.

2.3. Testing protocol and data collection

TA sessions were held in the physicians' actual workplace. After training the participants with the TA method in 10-min sessions, they were given five scenarios containing seven to nine tasks. The participants were provided with TA instructions and the clinical information for each scenario. A radiologist and an internist physician, both with at least one year of experience working with PACS, were consulted for the design of scenarios. These scenarios contained different modalities, including two digital radiography (DR), two computed tomography (CT), and one magnetic resonance imaging (MRI). Both the radiologist and the internist physicians used the same scenarios except for the MRI containing study, which was run only by the radiologist, as it was a brain tumor case and irrelevant to the internist physicians. The scenarios were designed in a manner to examine different parts and functions of a PACS and covered the most common tasks that a clinician may use in a typical working application. Generally, the scenarios included the following tasks: lesion size measurement, densitometry (Hounsfield unit), contrast change and window level, zoom, magnification, and layout use (observation of images in different cuts and views).

The scenarios and tasks were offered to the radiologists and internists. Table 1 illustrates the scenarios, goals, and actions needed to complete the tasks.

A CD containing the medical images of patients and consistent with the scenarios was played for the physicians during each TA session. The ID number of each image matched the scenario number, allowing the physicians to perform the tasks via case-by-case patient selection.

The clinical images of the patients used in the scenarios were collected from the hospital database. All patient-identifying information was deleted from the PACS images to maintain confidentiality. The study was reviewed and approved by the ethics committee of Kerman University of Medical Sciences (IR.kmu.REC.1394.454).

Interaction with the system was done through a standard mouse and a keyboard. Capturing the video, audio, and all of the activities on the

Table 1
Descriptions of the scenarios used in the usability test.

	Scenarios	Goals	Tasks	Actions
1	An abdominal CT was performed on a patient suspected of nephrolithiasis. Please check the pelvis in terms of tumors, cysts, and abnormalities by changing the contrast and window level.	Changing the image contrast and window level	1. Open the patient's image 2. Adjust and change the image contrast by dragging the cursor	1.1. Double click on the study 2.1. Find the windowing tool 2.2. Change the image contrasts by dragging the cursor on the image 2.3. View the image in the better window
2	A lung CT was performed on a patient hospitalized in infectious ward. Please determine the size of the lesion in the HRCT image of the patient.	Determining the size of the lesion	1. Open the patient's image 2. Determine the size of the lesion	1.1. Double click on the study 2.1. Scroll through the image to find the lesion 2.2. Find the measurement tool 2.3. Use the measurement tool by drawing horizontal and vertical lines
3	An abdominal CT was performed on a patient with a hepatic mass. Please check the density of the lesion using the magnifier.	1. Using the magnifier tool to find the lesion and determine its density. 2. Determining the density of the lesion	1.1. Open the patient's image 1.2. Enlarge the lesion using the magnifier tool 2.1. Determine the Hounsfield units	1.1.1. Double click on the study 1.2.1. Find the magnifier tool 1.2.2. Scroll the magnifier tool to find the lesion 2.1.1. Find the Hounsfield units tool 2.1.2. Report the number of the Hounsfield unit
4	A brain MRI scan was performed to evaluate tumor recurrence in a patient with a history of brain tumors. Please determine the lesion size in one layout.	1. Observing the lesion in a different layout 2. Determining the size of the lesion	1.1. Open the patient's image 1.2. Take the image in different layouts 2.1. Determine the size of the lesion in one layout	1.1.1. Double click on the study 1.2.1. Find the layout tool 1.2.2. Select one of the layouts 2.1.1. Find the measurement tool 2.1.2. Draw horizontal and vertical lines using the tool 2.1.3. Report the size of the lesion
5	A chest MRI was performed on a patient suspected of having a mass in the chest. Please check the size of the lesion with the zoom and magnifier tools.	1. Finding the lesion using the zooming tool 2. Observing the lesion using the magnifier tool 3. Determining the size of the lesion	1.1. Open the patient's image 1.2. Enlarge the entire image using the zooming tool 2.1. Enlarge the lesion using the magnifier tool 3.1. Determine the size of the lesion	1.1.1. Double click on the study 1.2.1. Find the zoom tool 1.2.2. Enlarge the entire image using the zooming tool 2.1.1. Scroll through the images to find the lesion 2.1.2. Enlarge the lesion using the magnifier tool 3.1.1. Find the measurement tool 3.1.2. Use the measurement tool by drawing horizontal and vertical lines 3.1.3. Report the size of the lesion

computer screen including the number of mouse clicks and keystrokes and elapsed time, was conducted using Morae recorder version 3.3 (TechSmith Corporation, Okemos, MI, USA). Each TA session lasted 20–40 min. The participants were asked to complete the PSSUQ immediately after the TA session.

The order of scenarios and PACS systems for each participant was altered and a 2-week time interval was planned between each of the participant's tests to increase the validity of the results and decrease the learning effect.

2.4. Measurement

Measurements were addressed based on a coding framework mentioned in the study design and procedure section. The usability characteristics and problems and their severity rating are described as follows:

Efficiency: It was measured by two metrics: (1) the number of mouse clicks and keystrokes and (2) the task completion time.
Effectiveness: It was measured by the number of completed tasks (task completion rate), which indicated the tasks' success rate.
Learnability: It was evaluated by measuring the number of tasks that were easily completed.
Errors: They were identified as the number of user mistakes when performing the tasks.
Satisfaction: It was measured by the participant's response mean score. Usability dissatisfaction problems were assessed through comments left by the users for each task or system option. The evaluators categorized some differences in the PACS systems by extracting the important points mentioned by the participants in the videos. These differences were categorized based on the tasks in the scenarios (measurements, zoom and magnifying, and contrast and window level).
Usability problems: They were detected based on the analysis of problems encountered by the participants during the interactions that were detected from the video reviews. In addition, the usability problems derived and detected from the coding framework were also considered.
Severity: It was measured via the mean severity rating determined by the evaluators.

2.5. Data analysis

A protocol analysis [30] was performed on all verbal utterances of the participants. All audio and video recordings of the TA sessions were reviewed by three evaluators to detect usability problems and the way the participant performed the task. The data were classified, analyzed, and coded based on the framework using the Morae manager.

Each evaluator was independently provided with a list of usability problems encountered by the participants. The collected data were merged into a unique master problem list and the disagreements about identified problems were discussed and resolved in evaluators' joint meetings by reviewing the video and audio data.

To analyze the results and categorize the identified problems, three formal joint meetings were held, each lasting approximately 1 h with short breaks. The evaluators independently determined the severity of each identified problem in each PACS based on the five-point rating scale proposed by Nielsen [37,38]. The problem severity rate was calculated based on a combination of three factors including the frequency of the problem, the potential impact of the problem on the user, and the persistence of the problem each time the user encountered the same situation [38]. Another meeting (30 min) was scheduled to summarize and calculate the average severity of the problems.

The data were analyzed using SPSS version 20 (SPSS Inc., Chicago, IL, USA). Each of the eight participants tested all three software programs. Given that each PACS in the study was tested several times by the participants, a linear mixed model test was used to investigate the differences of five usability characteristics between the PACS systems. The Bonferroni test was used for a pairwise comparison between the PCAS systems to investigate the provenance of discrepancy. The number and severity of the problems in each PACS did not affect the multiple measurements. Thus, the chi-square test was used to compare the number of problems in each PACS one-way ANOVA was used to compare the mean severity of the PACS systems.

3. Results

Eight participants were enrolled in the study; six were male. The participants' mean age was 45 years (ranging from 40 to 55 years).

Fig. 1 shows the details of usability characteristics for each PACS.

The comparison of results based on the five usability characteristics as well as on the number and severity of usability problems are as follows:

3.1. Efficiency

3.1.1. Task time

The average task performing time in PACS A was significantly less ($p = 0.01$) than that for other PACS systems (Fig. 1). The linear mixed model showed a significant difference in task times ($p = 0.01$) between the PACS systems. The Bonferroni test found that this difference was between PACS A and C ($p = 0.01$).

3.1.2. The number of mouse clicks and keystrokes

The linear mixed model test showed that PACS A required fewer mouse clicks and keystrokes than the other PACS systems (Fig. 1). However, the difference was not significant ($p = 0.12$).

3.2. Effectiveness

On average, the number of completed tasks in PACS B was larger (Fig. 1). The linear mixed model indicated a significant difference between the PACS systems ($p = 0.005$). The Bonferroni comparison showed that this difference was between PACS B and C ($p = 0.003$).

3.3. Learnability

The participants completed their tasks more easily in PACS A. There was a significant difference in the number of easily completed tasks between the PACS systems ($p = 0.001$). Bonferroni showed that this difference was between PACS A and C ($p = 0.001$) and PACS B and C ($p = 0.005$).

3.4. Error

No significant difference in the number of user errors was seen between the PACS systems ($p = 0.18$).

3.5. Satisfaction

As shown in Fig. 1 and based on the mean questionnaire scores, the participants assigned a greater satisfaction score (4.02) to PACS A. Also, a comparison between the mean participants' satisfaction responses showed a significant difference between PACS systems ($p = 0.009$). Examples of some differences (positive and negative aspects) between the PACS systems that were mentioned by the participants are shown in Table 2.

The number of problems and mean severity in each PACS are shown in Fig. 2. PACS A and C had 11 and 13 usability problems, respectively. The mean severity for each PACS was almost the same.

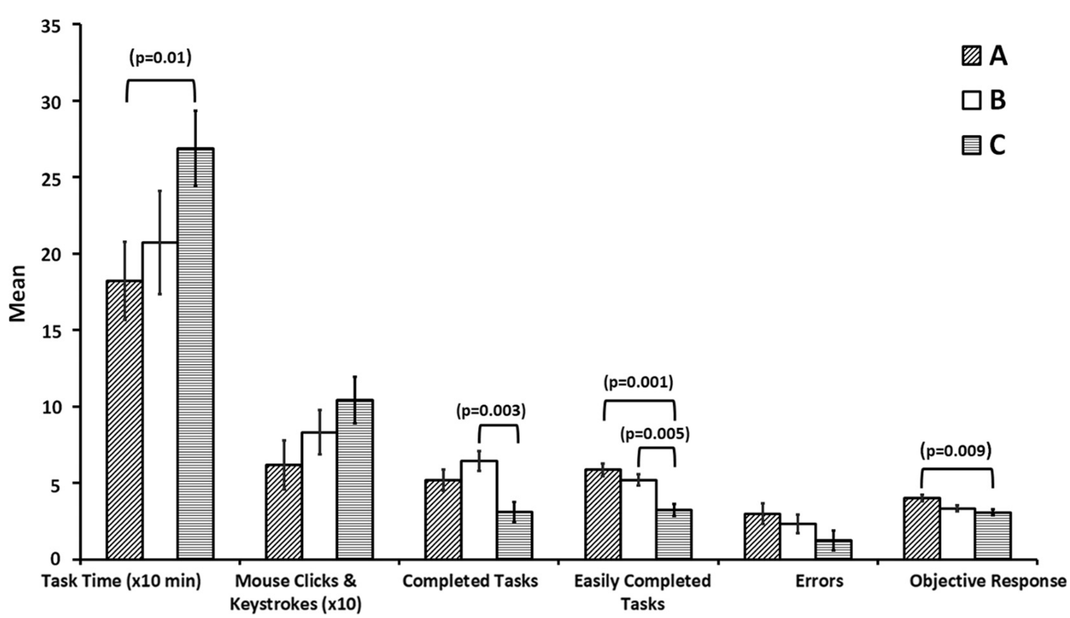

Fig. 1. Comparing measures of usability characteristics among A, B, and C PACS systems. The p values < 0.05 show significant differences.

Table 2
Differences in a sample of functionalities of the PACS systems.

Functionalities	PACS A	PACS B	PACS C
Measurement tool	Uses a narrow line that indicates the beginning and end of the lesion. (+)	Uses a thick line that covers the lesion borders. (−)	Uses a line that has a similar color to the pictures. (−)
Zoom and magnifying tool	Uses a big magnifier window that is active with one click and does not require repeated clicking of the mouse. (+)	Uses a smaller window and requires repeated clicking of the mouse. (−)	Does not focus on one part of a lesion and enlarges the whole picture. Does not use any magnifying window. (−)
Icons	Icons represent the corresponding functions. (+)	The zooming and magnifying icons are similar. (−)	The measurement icon does not look like a ruler. (−)

(+): Positive aspect. (−): Negative aspect.

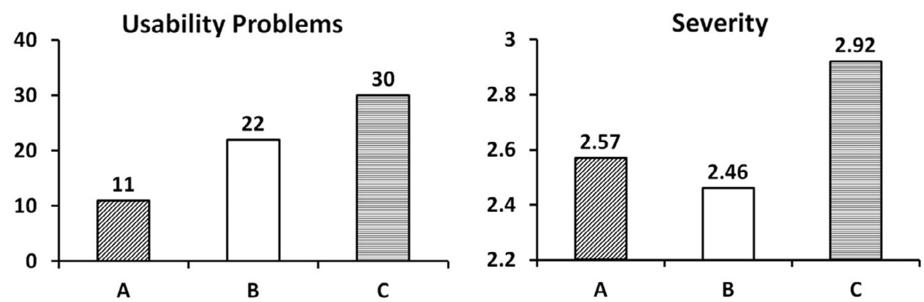

Fig. 2. The number of usability problems and the severity of problems for each PACS.

3.6. Usability problems

There was no significant difference in the total number of usability problems between PACS systems (p = 0.6). However, there was a borderline (marginal) significance in terms of severity (p = 0.09).

4. Discussion

Considering the users' easy interaction with the PACS systems, this study was designed to reveal the users' interaction problems and perspectives about three commonly used PACS systems. Our findings showed that applying the approach proposed in this study to augment TA usability evaluation (a real user's interaction) with user perspectives was able to identify a significant number of usability issues. The augmented method and the classification of the interaction problems based on the usability characteristics helped us select the best PACS that satisfied the users' real requirements. According to the results of the TA method, PACS A was found to be superior to the two other PACS systems in terms of efficiency, effectiveness, learnability, and satisfaction. In addition, the total number of problems in PACS A was fewer than the others. PACS A was selected for the Kerman University of Medical Sciences (KUMS) hospitals.

The think aloud analysis showed that the efficiency of PACS A was significantly greater than the two other PACS systems (B and C) (p = 0.01). This efficiency was more concerning the task time than the number of mouse clicks and keystrokes. There was a significant difference (p = 0.01) between the PACS systems in terms of the task time that can be attributed to the availability of tooltips. PACS A has appropriate (visible and self-explanatory) tooltips that help users find the options faster. Tooltips in a software program enable users to find the options easily via graphics and references [39]. When looking for a particular option, the users can understand the relevance by simply keeping the cursor under each item. Therefore, tooltips can lead to a reduction in the task time in PACS A. The slower performance of PACS B and C was caused by the illogical location of the tools in the menu, making them more difficult to find. This was the same reason for increased task time in PACS B and C in a similar study [18].

Our study found no significant difference between the three PACS in the number of mouse clicks and keystrokes as an indicator of efficiency (p = 0.12). However, the mean score of the number of mouse clicks and keystrokes in PACS A was less than the two other PACS systems. This may be due to the fact that in PACS A, one click was enough to choose a series of images, whereas the other PACS systems required double-clicking or cursor dragging. Also, options such as magnifier or densitometry (measuring the Hounsfield unit) for a lesion in PACS A did not need several mouse clicks. The large number of mouse clicks and keystrokes in PACS B and C that reduced the efficiency was related to using keyboard shortcuts. For example, the participants had to press control (Ctrl) and one of the F keys simultaneously to change the window level in PACS B and C. This usually causes difficulty for users and increases the number of mouse clicks and keystrokes. This issue was in contrast to the results of similar studies [40,41] that indicated keyboard shortcuts could be efficient and have the advantage of providing fast access to all commands. Choosing a series in PACS B and C was done by double-clicking or dragging and dropping, which could increase the number of mouse clicks. The cause of increasing the number of mouse clicks in our study differed from a similar study in which the participants had to switch between tabs to perform a particular task; thus, more actions were needed, leading to an increased number of mouse clicks and keystrokes [18]. In another study, Feizi [42] also realized that in a usable software application, fewer numbers of steps necessary to accomplish tasks are analogous to fewer numbers of mouse clicks and keystrokes.

The effectiveness of PACS A in our study was significantly (p = 0.005) more than the two other PACS systems (B and C). Concerning the task completion as an indicator of effectiveness, the results indicated that there was a significant difference between the PACS systems (p = 0.005). The number of completed tasks in PACS A was more than the other two PACS systems. This may be because the tasks in PACS A were in line (compatible) with the steps for doing a task, allowing the participants to complete more tasks. Assuming that the participants would be able to complete all tasks on all PACS systems, Jorritsma et al. [18] did not measure the task completion rate, but acknowledged that the measure of effectiveness, especially when more complex tasks are used in the test, might improve the accuracy of the usability assessment.

PACS A showed significantly higher (p = 0.001) learnability compared to the two other PACS systems. There was a significant difference between the PACS systems in the number of tasks completed with ease as an indicator of learnability (p = 0.001). It is easy for participants to learn PACS A. Users can find options easily in PACS A since its tooltips are larger and clearer. It also has appropriate icons that clearly represent the action of each option and help users understand the system more easily. Consistent with our study, a review on user interface design principles to increase software usability found that using larger components and bigger icons for key functions of the software can solve learnability related problems [43]. Improving graphical icons has also been shown to lead to improvements in mobile device learnability [44]. The difference in both task time and learnability was significant between the PACS systems in this study. A learnable system has been defined as a system that requires less time for the users to complete their tasks [45]. In line with the relevant literature [45,46], our results indicated that tasks are done in less time with more learnable PACS systems. Jorritsma et al. [18] did not directly investigate learnability; however, their selected PACS had less task completion time, which can be inferred as a more learnable task.

In contrast to studies on the selection of health systems that prefer low-error PACS [47], PACS A, as the selected system in our study, had a significantly higher number of errors compared to PACS B and C ($p = 0.18$). This is in line with Rodrigues et al.'s study [25] that indicated PACS systems, as critical and complex systems, have a number of problems that make them more error-prone and demanding of effort by users. No significant difference was found in the number of user mistakes as an indicator of error among different PACS systems ($p = 0.18$). Because choosing a wrong option increases the number of user mistakes, increasing the number of errors may be due to the unnecessary clicks; therefore, it cannot be inferred that our PACS systems were error prone. Unnecessary clicks occur when the participants click irrelevant buttons for searching the desired option without much thinking. This leads to an increase in the number of errors in all PACS systems.

The analysis of the user's perspective showed that the user's satisfaction for PACS A was significantly more than other PACS systems ($p = 0.009$). The results of the PSSUQ in this study showed a significant difference in the satisfaction scores of the PACS systems ($p = 0.009$). Generally, the users were satisfied with doing tasks in PACS A. Studies on the usability evaluation of health information systems have shown the importance of user satisfaction [10,48]. Consistent with the guidelines for designing user interface software, [49,50] the results of this study showed that using a screen with a larger display area for images and a smaller area for menus and toolbars increases the PACS user friendliness. The participants expressed a strong dislike for PACS B screen as it did not use a full screen to represent the images. They stated that despite having useful options, working with PACS B was complicated due to the disordered organization of information on the screen. This usability satisfaction issue was also noted in a study by Jorritsma et al. [18]. Another previous study [42] also reported that an organized interface arrangement in a usable software application led to the highest convenience as well as improvement in user satisfaction. According to another study [51], PACS software must contain a high precision measurement tool as a necessary requirement in its design. The participants in our study stated that PACS A was more accurate since the arrows could correctly determine the two sides of a lesion in an image. On the contrary, PACS B used an unsuitable thick line with no decimals for these measurements, leading to reduced accuracy. PACS C lacked measurement precision because of the difficulty in reading the numbers. This usability issue was mentioned in a prior study [42] indicating that ease of reading is important for user satisfaction with a software application. The isochromatic measurement line in PACS B and C caused a problem with reading the numbers (Table 2). The measurement tool was deactivated after each use when the participants had to make consecutive measurements; hence, reselection of the measurement tool was necessary each time. This dissatisfaction problem was reported in unselected PACS in a study by Jorritsma et al. [18]. This usability issue is in contrast to the results of another study by Jorritsma [17] that indicated measurement tools stay active until another tool is selected. In line with Gamma's study [52], the left or right side screen menus were more preferable compared to the top screen menus. The tiny black and white icons accumulated on the screen top with no toolbars and tabs as well as improper alignment of the options in PACS C dissatisfied the participants. The participants in Jorritsma et al.'s [18] study also preferred a toolbar consisting of multiple tabs, each with a different set of functionalities. According to the results of this study (Table 2), PACS systems should use large windows for zooming and magnification. PACS A had better functionalities because of the larger window. PACS A and B used windows for zooming in on each desired area of the images, while in PACS C the whole image was enlarged at once and users could not see the lesion in high resolution. In similar studies, the zooming tool was not provided in the interface [53] or it had an unexpected effect [54]. Our findings such as the examples provided in Table 2 suggest some features that should be considered in designing a PACS user interface. Using the full screen and suitable left or right side screen menus with familiar and clear labels make PACS more user-friendly. Providing functionalities such as an accurate and easy to use measurement tool and a large window for magnifying, zooming, and densitometry that can be activated with the least possible number of mouse clicks, icons that clearly represent the action of each option, and a suitable user-friendly toolbar design with ease of access would improve the usability of a PACS system.

Compared to the study by Jorritsma et al. [18] that recruited only four radiologists, in this study, we used two groups of PACS users, four radiologists and four internist physicians, to represent each group's needs. We recruited internist physicians as the second group of the PACS users, since in integrated RIS/PACS physicians also need and use PACS apart from radiologists [55,56]. The results indicated that except in a few cases, the radiologists' perspectives were the same as internist physicians'. So in applying the usability evaluation to the PACS selection process, if access to radiologists would be difficult, internist physicians could be used as an alternative sample. However, the few differences between radiologists and internist physicians' perspectives should be considered. These differences could be due to varying users' needs or different daily functions. Diagnosis is more important for radiologists and treatment is more important for internist physicians. The most obvious difference was regarding the precision of a tool, especially the measurement tool, which is more important for radiologists since they must report the exact size of a lesion. Similar to our study, Jorritsma et al. [18] compared four PACS systems with usability testing methods. They measured only the satisfaction and efficiency aspects of usability and ignored other characteristics. Overall, efficiency and satisfaction were considerably measured in previous usability evaluations of PACS studies [17,20,54], but other usability characteristics were not measured or measured less. For a more comprehensive and accurate usability evaluation, we additionally studied other usability characteristics including effectiveness, learnability, error, the number of usability problems, and the severity rating. Our results showed that in addition to user satisfaction and efficiency, which were emphasized in previous studies [17,20,54], other usability characteristics could substantially affect the stakeholders' decision to select a PACS. Hence, it is suggested to consider all usability characteristics when selecting a PACS, namely those proposed by ISO and Nielsen [33–35].

Unlike a former study [18] that used webcams with screen capture software, we used the TA method combined with Morae. Jorritsma et al. [17] determined the number, nature, and severity of usability issues that radiologists encountered in a different study but did not use the Nielsen severity rating. In fact, they did not use the list of usability problems and did not determine the severity of each problem.

In line with Nielsen's study [57] on the correlation between satisfaction and objective usability metrics, the qualitative user perspectives data obtained from the questionnaires in our study were consistent with the quantitative results of the TA method performance data.

Most evaluation studies on PACS used survey methods

[53,54,58–61]; hence, they did not directly evaluate users' interaction with systems. A major strength of our study is that the usability evaluation test directly examined end users' real interactions with the system and helped us more precisely evaluate objective metrics such as the efficiency of users' interactions. Although the time interval between the tests was reasonable, we altered the order of the scenarios and the PACS systems for each participant in order to obtain a higher validity of results and reduce the learning effects.

This study evaluated basic image manipulation tools such as zoom, contrast, and others. These tools were different from those in a study by Jorritsma et al. [18]. For example, we investigated the layout tool, which was not evaluated by Jorritsma et al. [18]. As another example, we did not investigate the 3D tool, while it was evaluated by Jorritsma et al. [18]. Considering that each tool in a PACS represents a specific function, investigating the user-friendliness of a PACS may depend on evaluating the usability of numerous tools or functionalities of that PACS. We need to design scenarios for evaluating the usability of tools or functionalities. As is known, in usability tests, scenarios were designed based on the needs of the user group (end users) [62]. In other words, the design of scenarios depends on the specialty of the participants using the system. Functionalities that were embedded in the scenarios in our study were also based on the common needs of internist physicians and radiologists. Thus, for evaluating all tools and functionalities of a PACS, participants from different user groups should be invited to usability tests. Having a scenario for each tool of a PACS, based on the needs of most different groups of end users, can increase the knowledge about the user-friendly design of that tool.

As described in this study and the results of a recent review article concerning usability problems in health information systems for radiology, PACS systems suffer from many usability problems. The results of the systematic review showed that the most common and critical usability problems in PACS systems concerned lack of flexibility and consistency of design [25]. The essence of the problems identified in our study is similar to the problems that the systematic review reported. Therefore, the future (re)design of PACS systems should follow design principles focusing on flexibility and consistency of user interfaces.

This study had four limitations. First, we did not pair the participants based on personal characteristics such as manual dexterity, which may negatively affect the data in this study. It is likely that sometimes increasing the task time may be due to factors related to a user's individual attributes rather than system usability problems. This factor did not have a major effect on this study since we used the same users for the tests. Faster participants with one PACS could perform quickly on the other two PACS systems as well. However, to entirely reduce this effect in addition to task time measurement, the number of mouse clicks and keystrokes, which are less affected by personal characteristics, were also quantified. Selecting participants with similar scientific knowledge levels and computer experience also reduced this potential effect. Second, due to the limited access to other PACS systems after selection, the memorability testing was not feasible in this study; however, we performed a comprehensive PACS usability evaluation through the integration of other usability characteristics proposed by ISO and Nielsen. Third, the current study considered usability problems and severity ratings as independent factors. Since the relevance of a particular usability problem in a PACS can be offset by another with a different severity, the comparison of systems based on independent factors should be done with caution. However, this risk was very low in this study because PACS A had the fewest usability problems and the lowest severity rating. We suggest that future studies use a weighted scale that indicates a better view of problems and their impacts on users [63]. Fourth, this study was conducted on three PACS systems with the participation of eight users. This might limit the generalizability of the results to other dissimilar systems. Although a relevant study suggest that only four to five individuals are sufficient to expose 80% of the usability issues in a given system [64], the eight participants used in this study might be insufficient for a robust quantitative analysis. Radiologists and internists are very busy with their clinical routines and have tight schedules; hence, it is difficult to recruit them. Moreover, it is often not cost-effective to engage more participants. However, compared to previous similar studies [53,54,65], we recruited the same or a higher number of participants in our study. Chen et al. [66] recommended overcoming the problem of the participation of a low number of specialist clinical users by selecting suitable usability techniques. This study augmented two relevant methods and analyzed and discussed qualitative indicators including the main usability characteristics proposed by ISO and Nielsen to improve the accuracy of the usability evaluation.

Proper selection of a usable, user-friendly, and user-satisfying PACS as a recently emerging and gradually expanding technology is important; thus, usability evaluation considering all possible characteristics can be helpful in decision making. In addition, applying user testing methods can be beneficial to compare various PACS options and functionalities.

As with any type of preventive measure, usability testing before system selection and implementation could potentially save money and time and improve the ultimate success of any PACS implementation project.

According to the obtained results, comprehensive usability evaluation considering all aspects of usability characteristics with all user groups is suggested as a criterion in PACS selection. This study has addressed some usability issues that should be noted in the design of PACS software. Also, the methodology can be used for any institutions selecting a PACS.

Measuring the memorability with repeated tests for participants with different medical specialties in each PACS is also suggested for future studies.

5. Conclusion

This study showed that applying the approach proposed herein to augment TA that focuses on a real user's interaction with user perspectives, addresses a wide range of theoretical aspects of usability, and can be beneficial for selecting the most usable PACS. Applying this technique during the vendor selection process enables purchasing organizations to compare features and items of PACS systems that best fit the qualifications and needs of their users, along with other selection criteria. Given that TA is the gold standard of usability evaluation [29], the results suggest that augmentation of the users' perspectives adds value to the results of TA. We suggest considering all usability characteristics when selecting a PACS, namely those proposed by ISO and Nielsen. Radiologists and clinicians encounter a large number and wide range of usability issues when using PACS systems in clinical practice; these should be addressed in the re-design of a PACS user interface. PACS providers should pay more attention to the usability of their products by applying such evaluation methods to improve their design and user interface.

Authors' contributions

M. Zahiri and R. Khajouei formed the concept and designed the study, acquired and interpreted the data, and drafted the paper. M. Baneshi had the main responsibility for the statistical analysis of the data. All three authors read and approved the final version of the submitted article.

Conflict of interest

The authors declare no conflicts of interest in this study.

Acknowledgments

We are grateful to the radiologists (Sarir Nazemi, Ahmad Enhesari, Farnaz Fahimi, and Amir Reza Salehi) and the internists (Seyed Mahdi Hashemi, Mohammad Reza Ebadzade, Syed Mahdi Seyedmirzaee, and Rostam Yazdani) for their participation in the usability test. We also thank Kambiz Bahaadinbeigy for facilitating communication with the participants.

References

[1] G. Bansal, Digital radiography. A comparison with modern conventional imaging, Postgrad. Med. J. 82 (2006) 425–428.
[2] U. Ewert, U. Zscherpel, K. Bavendiek, Replacement of Film Radiography by Digital Techniques and Enhancement of Image Quality, Annual Conference of Indian NDT Society, Kalkutta, 2005.
[3] R.H. Choplin, J. Boehme 2nd, C. Maynard, Picture archiving and communication systems: an overview, Radiographics. 12 (1992) 127–129.
[4] M.-N. Rohaya, Medical imaging trends and implementation: issues and challenges for developing countries, J. Health Inform. Dev. Ctries 5 (2011).
[5] P. Sylva, A situation analysis on PACS prospects for a developing nation, Sri Lanka J. Bio-Med. Inform. 1 (2010).
[6] S.A. Kim, W.-S. Park, T.J. Chun, C.M. Nam, Association of the implementation of PACS with hospital revenue, J. Digit. Imaging 15 (2002) 247–253.
[7] S. Ajami, R. Arab-Chadegani, Barriers to implement Electronic Health Records (EHRs), Mater Sociomed. 25 (2013) 213–215.
[8] M. Heydari, F. Saghafi, M. Khansari, Effective factors for implementing PACS in Iran regard to future technology trends, J. Med. Council Islamic Republic Iran 31 (2013) 265–271.
[9] K.G. Adler, How to select an electronic health record system, Fam. Pract. Manag. 12 (2005) 55.
[10] N.J. Corrao, A.G. Robinson, M.A. Swiernik, A. Naeim, Importance of testing for usability when selecting and implementing an electronic health or medical record system, J. Oncol. Pract. 6 (2010) 120–124.
[11] B.F. Branstetter IV, Basics of imaging informatics: part 1 1, Radiology 243 (2007) 656–667, http://dx.doi.org/10.1148/radiol.2433060243.
[12] D.L. Weiss, K.M. Siddiqui, J. Scopelliti, Radiologist assessment of PACS user interface devices, J. Am. Coll. Radiol. 3 (2006) 265–273.
[13] B. Bygstad, G. Ghinea, E. Brevik, Software development methods and usability: perspectives from a survey in the software industry in Norway, Interact. Comput. 20 (2008) 375–385.
[14] A. Kushniruk, Evaluation in the design of health information systems: application of approaches emerging from usability engineering, Comput. Biol. Med. 32 (2002) 141–149.
[15] A.W. Kushniruk, V.L. Patel, J.J. Cimino, Usability testing in medical informatics: cognitive approaches to evaluation of information systems and user interfaces, in: Proceedings of the AMIA Annual Fall Symposium, American Medical Informatics Association, 1997, p. 218.
[16] B. Aldosari, User acceptance of a picture archiving and communication system (PACS) in a Saudi Arabian hospital radiology department, BMC Med. Inf. Decis. Making 12–44 (2012).
[17] W. Jorritsma, F. Cnossen, R.A. Dierckx, M. Oudkerk, P.M. Van Ooijen, Post-deployment usability evaluation of a radiology workstation, Int. J. Med. Inform. 85 (2016) 28–35.
[18] W. Jorritsma, F. Cnossen, P.M. van Ooijen, Merits of usability testing for PACS selection, Int. J. Med. Inform. 83 (2014) 27–36.
[19] S.-C. Wu, J. Smith, J. Swan 2nd, Pilot study on the effects of a computer-based medical image system, in: Proceedings of the AMIA Annual Fall Symposium, American Medical Informatics Association, 1996, p. 674.
[20] W. Jorritsma, F. Cnossen, R.A. Dierckx, M. Oudkerk, P.M. van Ooijen, Pattern mining of user interaction logs for a post-deployment usability evaluation of a radiology PACS client, Int. J. Med. Inform. 85 (2016) 36–42.
[21] M. Chow, L. Chan, Development and evaluation of a compartmental picture archiving and communications system model for integration and visualization of multidisciplinary biomedical data to facilitate student learning in an integrative health clinic, Comput. Educ. 54 (2010) 733–741.
[22] B.J. Erickson, M.F. Kossack, Application of usability testing for improving PACS workstation design, in: Medical Imaging 2000, International Society for Optics and Photonics, 2000, pp. 188–195.
[23] K. Inamura, J. Konishi, H. Nishitani, S. Kousaka, Y. Matsumura, H. Takeda, et al., Status of PACS and technology assessment in Japan, Comput. Methods Progr. Biomed. 66 (2001) 5–15.
[24] S. Zakerabasali, H. Monem, R. Sharifian, R. Khajouei, Usability evaluation of Ulite web-based picture archiving and communication systems software in Shiraz University of medical science hospitals, J. Health Manage. Inform. 2 (2015) 132–137.
[25] C. Rodrigues Dias, M. Rodrigues Pereira, Freire A. Pimenta, Qualitative review of usability problems in Health Information Systems for radiology, J. Biomed. Inform. (2017).
[26] S. Carmien, M. Dawe, G. Fischer, A. Gorman, A. Kintsch, J.F. Sullivan Jr., Socio-technical environments supporting people with cognitive disabilities using public transportation, ACM Trans. Comput. Hum. Interact. 12 (2005) 233–262.
[27] N. Bazak, G. Stamm, F. Caldarone, J. Lotz, A. Leppert, M. Galanski, PACS Workstations 2000: Evaluation, Usability and Performance, EuroPACS, Citeseer, 2000, pp. 133–142.
[28] T. Boehm, O. Handgraetinger, J. Link, R. Ploner, D.R. Voellmy, B. Marincek, et al., Evaluation of radiological workstations and web-browser-based image distribution clients for a PACS project in hands-on workshops, Eur. Radiol. 14 (2004) 908–914.
[29] H.R. Hartson, T.S. Andre, R.C. Williges, Criteria for evaluating usability evaluation methods, Int. J. Hum. Comput. Interact. 13 (2001) 373–410.
[30] K.A. Ericsson, H.A. Simon, Protocol Analysis: Verbal Reports as Data, Rev. ed., MIT Press, Cambridge, MA, 1993.
[31] A. Holzinger, Usability engineering methods for software developers, Commun. ACM 48 (2005) 71–74.
[32] J.R. Lewis, IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use, Int. J. Hum. Comput. Interact. 7 (1995) 57–78.
[33] A. Abran, A. Khelifi, W. Suryn, A. Seffah, Usability meanings and interpretations in ISO standards, Software Qual. J. 11 (2003) 325–338.
[34] N. Bevan, Usability, in: L. Liu, M.T. ÖZsu (Eds.), Encyclopedia of Database Systems, Springer US, Boston, MA, 2009, pp. 3247–3251.
[35] J. Nielsen, Usability 101: Introduction to Usability, 2003. >https://www.nngroup.com/articles/usability-101-introduction-to-usability/> (accessed 12.06.17).
[36] T. Clemmensen, M. Hertzum, K. Hornbæk, Q. Shi, P. Yammiyavar, Cultural cognition in the thinking-aloud method for usability evaluation, in: ICIS 2008 Proceedings, 2008, p. 189.
[37] M.W. Van Someren, Y.F. Barnard, J.A. Sandberg, The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes, Academic Press, London, 1994.
[38] J. Nielsen, Severity Ratings for Usability Problems, 1995. <https://www.nngroup.com/articles/how-to-rate-the-severity-of-usability-problems/> (accessed 12.06.17).
[39] C. Knight, K. Mobley, T. Meserth, Tooltip Hyperlinks, Google Patents, 2002. <https://www.google.com.na/patents/US20040004632/> (accessed 12.06.17).
[40] D.M. Lane, H.A. Napier, S.C. Peres, A. Sandor, Hidden costs of graphical user interfaces: failure to make the transition from menus and icon toolbars to keyboard shortcuts, Int. J. Hum. Comput. Interact. 18 (2005) 133–144.
[41] R.C. Omanson, C.S. Miller, E. Young, D. Schwantes, Comparison of mouse and keyboard efficiency, in: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Sage Publications, 2010, pp. 600–604.
[42] A. Feizi, W. Chui Yin, Usability of user interface styles for learning a graphical software application, in: International Conference on Computer & Information Science (ICCIS), 2012, pp. 1089–1094.
[43] A. Darejeh, D. Singh, A review on user interface design principles to increase software usability for users with less computer literacy, Peer J. Comput. Sci. 9 (2013) 1443.
[44] R. Leung, J. McGrenere, P. Graf, The learnability of mobile application interfaces needs improvement, in: Proc. of British HCI Workshop on HCI and the Older Population, 2008.
[45] J. Sauro, How To Measure Learnability, 2013. <https://measuringu.com/measure-learnability/> (accessed 12.06.17).
[46] M. Laakkonen, Learnability makes things click: a grounded theory approach to the software product evaluation, Doctoral dissertation, University of Lapland, 2007.
[47] S. Bowman, Impact of electronic health record systems on information integrity: quality and safety implications, Perspect. Health Inf. Manag. 10 (2013) 1c.
[48] J. Palmius, Criteria for measuring and comparing information systems, in: Proceedings of the 30th Information Systems Research Seminar in Scandinavia IRIS, 2007.
[49] A. Monk, L. Davenport, J. Haber, P. Wright, Improving your Human-computer Interface: A Practical Technique, Prentice Hall, London, 1993.
[50] S.L. Smith, J.N. Mosier, Guidelines for Designing User Interface Software, Mitre Corporation, Bedford, MA, 1986.
[51] D.R. Haynor, D. Smith, H. Park, Y. Kim, Hardware and software requirements for a picture archiving and communication system's diagnostic workstations, J. Digit. Imaging 5 (1992) 107–117.
[52] E. Gamma, Design Patterns: Elements of Reusable Object-oriented Software, Pearson Education, India, 1995.
[53] D. Markonis, F. Baroz, R. Ruiz de Castaneda, C. Boyer, H. Müller, User tests for assessing a medical image retrieval system: a pilot study, Stud. Health Technol. Inform. 192 (2013) 224–228.
[54] I. Geldermann, C. Grouls, C. Kuhl, T.M. Deserno, C. Spreckelsen, Black box integration of computer-aided diagnosis into PACS deserves a second chance: results of a usability study concerning bone age assessment, J. Digit. Imaging 26 (2013) 698–708.
[55] S.H. Becker, R.L. Arenson, Costs and benefits of picture archiving and communication systems, J. Am. Med. Inform. Assoc. 1 (1994) 361.
[56] J. Nielsen, How Many Test Users in a Usability Study?, 2012. <https://www.nngroup.com/articles/how-many-test-users/ > (accessed 12.06.17).
[57] J. Nielsen, User Satisfaction vs. Performance Metrics, 2012. <https://www.nngroup.com/articles/satisfaction-vs-performance-metrics/> (accessed 12.06.17).
[58] R.A. Bauman, G. Gell, Reality of picture archiving and communication systems (PACS): a survey, J. Digit. Imaging 13 (2000) 157–169.
[59] S. Bryan, G.C. Weatherburn, J.R. Watkins, M.J. Buxton, The benefits of hospital-wide picture archiving and communication systems: a survey of clinical users of radiology services, Br. J. Radiol. 72 (1999) 469–478.
[60] D. MacDonald, D. Neville, R. Alaghehbandan, Referring physicians' perceived benefits of PACS: a regional survey pre- and post-PACS implementation, Radiol. Manage. 33 (2011) 46–54.
[61] B.I. Reiner, E.L. Siegel, F. Hooper, Z. Protopapas, Impact of filmless imaging on the frequency of clinician review of radiology images, J. Digit. Imaging 11 (1998)

149–150.
[62] M.B. Rosson, J.M. Carroll, Usability Engineering: Scenario-based Development of Human-computer Interaction, Morgan Kaufmann, 2002.
[63] R. Khajouei, A. Hasman, M.W. Jaspers, Determination of the effectiveness of two methods for usability evaluation using a CPOE medication ordering system, Int. J. Med. Inform. 80 (2011) 341–350.
[64] T.J. Robinson, S. DuVall, R. Wiggins, Creation and usability testing of a web-based pre-scanning radiology patient safety and history questionnaire set, J. Digit. Imaging 22 (2009) 641–647.
[65] D. Kim, P. Kang, J. Yun, S.-H. Park, J.-W. Seo, P. Park, Study on user interface of pathology picture archiving and communication system, Healthc Inform. Res. 20 (2014) 45–51.
[66] C. Chen, J. Abdelnour-Nocera, S. Wells, N. Pan, Usability practice in medical imaging application development, HCI and Usability for e-Inclusion, 2009, pp. 405–415.